IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 28, 2019, accepted August 4, 2019, date of publication August 15, 2019, date of current version September 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2935540

Towards Mobile Cognitive Fatigue Assessment as Indicated by Physical, Social, Environmental, and Emotional Factors

EDWARD PRICE[1], **GEORGE MOORE**[1], **LEO GALWAY**[1], **AND MARK LINDEN**[2]
[1]School of Computing, Ulster University, Newtownabbey BT37 0QB, U.K.
[2]School of Nursing and Midwifery, Queen's University Belfast, Belfast BT9 7BL, U.K.

Corresponding author: George Moore (g.moore@ulster.ac.uk)

This work was supported by the Department for Learning and Employment Northern Ireland and the School of Computing, Ulster University.

ABSTRACT This research sought to establish which in-situ measures of cognitive fatigue, physical activity, social interaction, location, emotional state and facial landmarks, made using a smartphone application, could be used to indicate episodes of cognitive fatigue. This assessment was realised using cognitive tests (assessing memory, attention, reaction time, information processing speed and executive function), self-assessment, contextual factors and facial feature analysis. This study also investigated the use of an ensemble algorithm for the classification of cognitive fatigue utilising facial features and a Rotation Forest approach. Self-assessment of cognitive fatigue was shown to directly correlate with reaction time through a Psychomotor Vigilance Task ($r = .643$, $p = .001$), and self-reported increases in the level of social activity ($r = .377$, $p = .001$). Facial feature analysis revealed dominant emotions of sadness and anger when participants were cognitively fatigued. It also revealed underlying facial cues that indicated higher levels of cognitive fatigue including expressions of negative valence, and Facial Action Coding System units of increased brow furrow, eyelid tightening and lip suck. In addition, a Principle Component Analysis based Rotation Forest ensemble with a ternary output demonstrated a cognitive fatigue classification accuracy of 82.17%. The findings presented indicate that the inclusion of data relating to surrounding cognitive, social, physical and emotional factors can improve the accuracy of mobile in-situ cognitive fatigue assessment using our previously validated smartphone-based cognitive fatigue assessment approach. The findings further suggest gross-level fatigue status may be potentially classified to a reasonable degree of accuracy using facial features, which may give rise to personalised in-situ fatigue detection.

INDEX TERMS Affective computing, cognitive fatigue, cognitive tests, context, facial analysis, human computer interaction, machine learning, mobile applications, neuropsychology, smart healthcare.

I. INTRODUCTION

Cognitive fatigue can be severe and debilitating and has been identified by people suffering from a range of conditions including Parkinson's Disease [1], stroke [2], heart failure and Acquired Brain Injury [3]. Cognitive fatigue can be caused by a lack of sleep [4], stress, or a cognitive deficiency [5] and can be triggered by carrying out simple activities of daily living. Diagnosis and assessment of problematic cognitive fatigue commonly relies upon periodic assessment that is conducted within a clinical setting. Predominantly, the traditional method of evaluating cognitive fatigue is self-assessment through questionnaires [6], [7], which primarily takes place under clinical supervision. A limitation of administering such assessments within a clinical setting is the need for a clinician to supervise, which is costly and time consuming. It also does not allow assessment in-situ, within the daily locations and routines of the patient, which is where problematic instances of cognitive fatigue occur. Development of easier to administer single-item scales has been proven as a way of assessing cognitive fatigue more efficiently than traditional clinic-based approaches [8]–[11]. Studies using state-of-the-art approaches have the possibility of utilising advances in technology, namely smartphone technology, facial feature analysis, and cognitive testing to provide a more easily accessible means of assessment [12]–[15].

The associate editor coordinating the review of this article and approving it for publication was Kaigui Bian.

The research presented in this paper sits within the broader goal of addressing the limitations of current in-situ cognitive fatigue assessment practices. This is done with a view to providing an in-situ evaluation technique that can be used to aid the assessment of cognitive fatigue and to help inform the extrinsic clinical evaluation of a person's condition. However, at this stage, the approach under investigation has a broader applicability in the assessment of cognitive fatigue, with a view to validating an approach that could later be tailored to the specific needs of the individual.

Building on our previous work, that showed the validity of assessing cognitive fatigue using game-like assessment tasks presented on a smartphone [16]–[19], this study aimed to establish which in-situ measures of cognitive fatigue, physical activity, social interaction, location, emotional state and facial landmarks, made using a bespoke smartphone application, could be used to indicate episodes of cognitive fatigue, as measured using our previously validated approach.

It was thought that by including surrounding factors, as well as a direct assessment of cognitive fatigue, it might be possible to gain a better understanding of a participant and their daily activities. This could then be used to refine the assessment of cognitive fatigue and to tailor the nature and timing of assessments in order to provide a more adaptive approach, which uses a minimal set of measures, while retaining the same degree of assessment reliability.

This study started from the understanding that it has already been established that a person's level of cognitive function can be measured using a range of appropriate cognitive tests [20], and that employing such an approach to assess cognitive fatigue using a smartphone-based assessment tool has also been shown to be valid in our previous work. This prior research investigated the relationship between validated self-assessment scales and cognitive tests presented on a smartphone and showed that both approaches are effective for measuring cognitive fatigue. The results from the current study will be assessed through comparison with results attained using the earlier validated approach.

As such, the research presented in this paper employed both subjective and objective measures to help to assess a participant's level of cognitive fatigue. Subjective self-assessment took place through the use of on-screen questions aimed at determining a participant's perceived cognitive fatigue, social interaction, location, and emotional state. Objective measures of current cognitive function were assessed, as an indicator of cognitive fatigue, using game-like tasks, namely: Spatial Span Task, Psychomotor Vigilance Task and a Mental Arithmetic Test; along with an objective measure of physical activity based on daily steps taken. These specific testing methods were selected as we have previously shown that they can measure a range of cognitive abilities including memory, attention, reaction time, information processing speed and executive function. Additionally, the smartphone assessment tool also captured a set of facial landmarks during each session, which were later analysed to determine facial cues and emotions determined to have been expressed during the assessment session. Such cues and emotions were based upon the Facial Action Coding System (FACS) developed by Paul Ekman [21]. This aspect was included to determine if there was any correlation between these objective measures of possible emotional state and those of cognitive fatigue made by the subjective and objective measures within the tool, which could lead to an alternative form of objective assessment. Furthermore, the facial landmarks were additionally utilised in the generation of an ensemble-based classifier with the goal of mapping facial feature data into one of three groups: low, normal and high levels of cognitive fatigue. Although offline within the study reported herein, the possibility of an efficacious classifier approach could potentially lead to further online modelling and the eventual generation of a real-time, personalised model of fatigue detection.

The study presented in this paper aimed to investigate the extent to which surrounding factors related to physical activity, social interaction, location, and emotional state could be measured and help to further assist in the assessment of cognitive fatigue. The efficacy of these measures was evaluated through comparison to results from subjective and objective measures of cognitive fatigue that were captured during the same testing session using an already validated approach. Correspondingly, a direct positive correlation was found between smartphone-based self-assessment of cognitive fatigue using a single-item scale and reaction time from the Psychomotor Vigilance Task ($r = .643$, $p = .001$), and between the self-assessed fatigue and self-assessed levels of social interaction ($r = .377$, $p = .001$). In addition, the analysis of facial features indicated a prevalence of the emotions of sadness and anger when participants self-reported cognitive fatigue, along with exhibiting several underlying facial cues that indicate higher levels of cognitive fatigue, such as expressions of negative valence, increased brow furrow, increased eyelid tightening and increased lip suck. A secondary aim of the work was to ascertain the potential of employing a machine learning approach for the detection of gross-levels of fatigue using facial landmarks obtained as part of the objective measures of emotional state. Subsequently, an ensemble-based classification approach was investigated using a range of classifiers, with a Principle Component Analysis based Rotation Forest ensemble achieving the highest classification accuracy (82.17%) when mapping facial landmarks to average reaction time from the Psychomotor Vigilance Task grouped into *low*, *normal* and *high* levels of cognitive fatigue.

It has been shown that the inclusion of surrounding factors can potentially improve the assessment of cognitive fatigue. This was particularly the case for self-declared location and some aspects of facial feature analysis. While this could, in turn, allow for a more adaptive approach to be taken, the results would need to be further validated within a larger longitudinal study. Additionally, the offline use of a Principle Component Analysis based Rotation Forest ensemble has been shown to be capable of classifying a low, normal or high level of cognitive fatigue from a set of facial landmarks with a reasonable degree of accuracy. Data from a longitudinal study would also facilitate further optimisation of the classification approach to yield improved accuracy, along with consideration of modern, online machine learning tools.

II. BACKGROUND

As noted, assessment of cognitive fatigue is not new and there are well established tools for use during assessment taking a traditional clinician-led approach. These generally involve the use of questionnaires or cognitive tests, often delivered by a medical professional. There are also existing technology-based approaches to assessing cognitive fatigue which include the simple adaption of traditional methods so that they can be delivered using technology, as well as new technology-based approaches that make use of sensor data. A range of both traditional and technology-based approaches are reviewed below, along with a consideration of their suitability for delivery within an in-situ patient-led approach, delivered using a smartphone; such as that taken within the study reported in this paper. Likewise, previous works related the use of machine learning approaches employed within fatigue detection and facial feature analysis are briefly reviewed.

A. TRADITIONAL APPROACHES TO COGNITIVE FATIGUE ASSESSMENT

Traditionally cognitive fatigue has been assessed through a clinician-led approach within a clinical setting. Such an approach typically makes use of some form of self-assessment to inform a clinical assessment. The tools used can include questionnaires, short-form questionnaires, or a variety of cognitive tests.

1) SELF-ASSESSMENT QUESTIONNAIRES

There are a number of self-assessment questionnaires designed to specifically assess cognitive ability and its relationship to fatigue, including the Mental Fatigue Scale (MFS) [22], Fatigue Severity Scale (FSS) [23] and Visual Analogue Scale for Fatigue (VAS-F) [24]. These scales use a visual analogue representation or targeted questions to aid a participant in self-evaluation. The MFS is a multidimensional questionnaire containing 15 questions developed by Johansson and Rönnbäck [25] and adapted from work by Rödholm *et al.* [26]. Similarly, the FSS is a nine-item questionnaire that assesses the effect of general fatigue on daily living where each item is rated on a 7-point Likert scale [23]. By contrast, the VAS-F, developed by Lee *et al.* [27], employs an 18-item visual analogue scale to allow participants to determine their own rating in regard to each statement presented in the scale. Typically, the questions presented within these scales measure effects such as: fatigue in general, lack of initiative, mental fatigue, mental recovery, concentration difficulties, memory problems and slowness of thinking. The drawback of these types of scales is that they are predominantly delivered within a clinical environment and can be long and somewhat difficult to complete, hence are not suitable for in-situ self-assessment.

2) SHORT-FORM SELF-ASSESSMENT QUESTIONNAIRES

In an effort to reduce the time on task and increase engagement, a number of single-item scales for the assessment of cognitive fatigue have been developed [8]–[11]. A concern with short-form questionnaires is that they might not fully assess the condition under consideration. However, the efficacy of these shorter self-assessment scales has been shown to be congruent with longer questionnaires [10], [28]. Furthermore, they are more manageable for a patient to complete, which in-turn makes them more suitable for in-situ patient-led assessment.

3) COGNITIVE TESTING METHODS

Cognitive testing methods have traditionally been used to determine cognitive ability; however, they can also be used to show discrepancies in performance relating to cognitive fatigue [29], [30]. The Psychomotor Vigilance Task (PVT) can be used to assess attention, speed of processing, and reaction time [31]; as such, it has been used as an accurate and valid measure of cognitive fatigue [32], [33]. Short versions of the PVT have also been evaluated and have been shown to be both successful and accurate at predicting cognitive fatigue [34]. In a study carried out by Johansson *et al.* [29], the ability of neuropsychological tests to measure cognitive fatigue was compared to the MFS in order to investigate potential correlations between objective and subjective assessment measures of cognitive fatigue. The study was carried out in a traditional clinical environment. Neuropsychological tests employed included: Digit Symbol Coding, Digit Span, Spatial Span [35] and Trail Making [36], which were used to measure processing speed, cognitive attention and working memory. Verbal fluency was measured though the FAS test, which requests an individual to orally produce as many words as possible that begin with the letter F, A and S within a prescribed timeframe [37]. Reading speed was measured using the DLS reading speed test. This test requires participants to read text in which three words appear at intervals within brackets. They must choose which of the words relates to the context of the sentence [38]. This study by Johansson *et al*. [29] established that subjective cognitive fatigue correlated with objectively measured information processing speed. The nature of these cognitive tests can lend themselves well to adaption within a technology-based assessment approach.

B. TECHNOLOGY-BASED COGNITIVE FATIGUE ASSESSMENT

Many traditional approaches to cognitive fatigue assessment have been adapted for use within a technology-based approach, and increasingly for delivery using a smartphone, thereby facilitating mobile assessment [39], [40]. In addition, new approaches to cognitive fatigue assessment have been facilitated through the use of sensors such as accelerometers for activity analysis and cameras for face analysis [20], [41].

1) ADAPTATION OF TRADITIONAL COGNITIVE FATIGUE ASSESSMENT METHODS

A number of studies into computerised cognitive fatigue assessment have looked to adapt traditional cognitive testing methods within a technology-based approach. While cognitive testing has traditionally been used to determine relative cognitive ability, it has also been shown to indicate discrepancies in performance in relation to cognitive fatigue [29], [30]. Originally designed for a static, desktop computer-based evaluation, the PVT has since been modified for use on a mobile phone in order to improve the utility of on-the-go assessment [30], [42]. The PVT has been shown to be an accurate predictor of vigilance due to fatigue and sleep loss, which can be a direct predictor of cognitive fatigue [32], [33]. Short-form versions of the PVT have also been shown to be successful in evaluating cognitive fatigue [34]. Work by Kay *et al.* [30], and Gartenberg and McGarry [42], investigated the efficiency of using short mobile-based tests, along with potential usability issues that may arise from test delivery using a mobile platform. Their work concluded that mobile-based assessment approaches were as effective as desktop computer-based approaches. Indeed, mobile phone-based versions of traditional assessment tools have more generally been shown to be as reliable in assessment as their paper-based counterparts, while at the same time making daily assessment easier to facilitate by merit of being available on a mobile phone [43], [44]. This supports the reliability of using cognitive tests within a smartphone-based cognitive fatigue assessment approach. However, when adapting existing assessment tools for delivery using a mobile phone, care must be taken to ensure that the tools are as simple to understand and complete as the corresponding paper-based tools they are replacing to avoid any usability issues effecting the efficacy of the tool [44].

2) PHYSICAL ACTIVITY RECOGNITION

Modern mobile devices have access to a multitude of sensors, which allow for a wide range of measures to be taken. This includes access to accelerometer, gyroscope and GPS sensors that can calculate physical activity and movement. Prescribed physical activity has long been known to help alleviate cognitive fatigue caused by multiple medical conditions [45]–[47]. Moreover, physical activity has not only been shown to aid rehabilitation but can also prevent cognitive decline and cognitive fatigue in the long-term [48], [49]. As such, a number of studies have investigated the use of physical activity tracking as a measure, and as a predictor, of cognitive fatigue. Accordingly, rather than prescribing a physical activity regime, simple analysis of daily activity levels could perhaps give a greater insight into daily cognitive fatigue levels and help inform rehabilitation and feedback [20], [50].

3) FACIAL ANALYSIS

The ubiquitous nature of high-quality front-facing cameras on smartphones gives good potential for the capture and analysis of facial images of a smartphone user. Facial feature analysis has been shown to be an effective measure of both cognitive and physical fatigue in specific cases, e.g. in assessing the fatigue of drivers [51], [52]. These studies showed that analysis of metrics such as head tilt, eyebrow position, mouth position and other facial cues can predict fatigue. A measure of the percentage of eye closure over a defined period of time (PERCLOS) has been determined as one of the most accurate ways of detecting fatigue via facial feature analysis [53]. Batista [41] and Gan *et al.* [51] carried out studies into the viability of using PERCLOS in order to detect both cognitive and physical fatigue, with results indicating that it was both a viable and accurate approach with regard to fatigue assessment. Similarly, the Driver Fatigue Monitor [54] is a real-time, video-based system that measures the degree of drowsiness of a driver, which employs a camera to detect slow eyelid closure and PERCLOS estimation, alongside head pose, to infer driver fatigue. It was reported from a study involving the Driver Fatigue Monitor that the metrics used could accurately indicate fatigue. However, care must be taken as many of these facial analysis studies are carried out in constrained contexts focusing on a single real-world use case. Due to this, the context, modality and effectiveness of a given approach may not directly translate to other situations such as that addressed in the study reported in this paper. However, there would seem to be some merit in exploring the potential of such an approach further, if not as a core objective of this study.

4) EMPLOYING MACHINE LEARNING APPROACHES

Research has also sought to employ machine learning approaches to cognitive fatigue detection and assessment, and facial feature analysis. Bundele *et al*. approached fatigue detection through analysis of skin conductance, respiration and blood oxygen levels [55]. Their implementation incorporated biometric skin sensors that relayed information to a smartphone for later analysis and evaluation though the machine learning approach of an Artificial Neural Network. Research carried out by Al-Libawy *et al*. [56] attempted to create a pervasive fatigue analysis system that did not require any additional sensors. Their approach analysed the primary input method of a smartphone (i.e. text entry), assessing the error rate as a psychomotor measure of fatigue. In research conducted by Kawamura *et al*. the accuracy and feasibility of a desktop-based, multi-camera system to capture facial features during speech for fatigue analysis was explored, with the authors reporting the use of a Support Vector Machine achieved a classification accuracy of 89% [15]. A Support Vector Machine was also utilised to classify the fatigue status of participants with a reported accuracy of 88.8%. Ghimire and Lee implemented Bagging for facial recognition [57] and used the Histogram of Orientation Gradient features [58] for facial feature analysis. Subsequently, they employed Bagging as the classification technique as it is more suited for use with the noisy data acquired. Fanelli *et al*. proposed a system for real-time 3D face analysis by applying a Random Forest to depth images, pose estimation and 3D facial features [59]. The system was capable of handling partial obstructions, along with noisy depth data that is acquired using everyday 3D sensors. Given the proliferation and widespread adoption of smartphone cameras, a potentially plausible method of passive fatigue analysis is the classification of facial features. Consequently, the quantity of complex data typically acquired introduces its own problems for feature selection and analysis. However, as briefly indicated, the use of classifier ensemble approaches such as Bagging [60], Boosting [61], Random Forest [62] or Rotation Forest [63] may provide a potentially plausible method for cognitive fatigue analysis through the classification of facial features. Smartphone-based cognitive fatigue assessment has been shown to have potential as a delivery platform for tools that assess cognitive fatigue in-situ, set within a person's activities of daily living. Such an approach has potential for delivering both revised versions of traditional cognitive fatigue assessment approaches, while retaining validity of measurement, and assessment tools that are technology-based by their nature, which have been shown to be able to measure cognitive fatigue. Moreover, the integration of machine learning-based analysis may provide additional tools to facilitate more robust, personalised fatigue detection and assessment.

III. OBJECTIVES

Building on the understanding developed during the background review, and in our previously published work [16]–[19] — in which we validated an approach to the in-situ assessment of cognitive fatigue using a set of game-like tasks delivered using a smartphone — this study aimed to address the research question: which in-situ measures of physical activity, social interaction, location, emotional state and facial landmarks, made using a bespoke smartphone application, could be used to indicate episodes of cognitive fatigue, as measured using our previously validated approach?

To address the above research question, this study tested the hypothesis that: measures of a participant's physical activity, social interaction, location and emotional state would be shown to be indicative of episodes of cognitive fatigue, as measured using our previously validated approach. It was further hypothesised that: facial feature analysis would allow for the selection of a machine learning approach capable of identifying episodes of cognitive fatigue.

Specifically, the factors considered against the objective measure of cognitive fatigue, obtained using our previously validated approach, were: i. subjective self-reported measure of cognitive fatigue, ii. objective measure of physical activity, iii. subjective self-reported recent social interaction level, iv. subjective self-reported location type, v. subjective self-reported emotional state, vi. objective emotional state, vii. objective facial expressions, viii. objective facial landmarks.

IV. METHODOLOGY

With the above hypotheses and measures in mind, a study was designed that would make use of a single bespoke smartphone application to gather data that could be used to test the hypotheses by delivering all required subjective and objectives measures within each test session.

A. SMARTPHONE-BASED COGNITIVE FATIGUE ASSESSMENT APPLICATION

The smartphone application used during this study was a revised and extended version of the application used during our earlier work to validate game-like tasks as a method of assessing cognitive fatigue on a smartphone. While the game-like tasks remained unchanged in this version of the application, the MFS questionnaire was removed. This step was taken as results from our previous study showed that our game-like tasks could reliably measure cognitive fatigue to a level in keeping with the MFS. Moreover, the same work showed that inclusion of the MFS in the smartphone application lead to reduced protocol adherence, largely due to the lengthy nature of the MFS questionnaire. Given the above points, it was decided that it would be appropriate to remove the MFS component of the original application, in favour of retaining only the validated game-like tasks as a baseline measure of cognitive fatigue.

For the purposes of addressing the aim of this current study, additional components were added to the smartphone application in order to help to determine factors surrounding completion of the game-like tasks. Single-item scales were added to the application to allow participants to subjectively self-report their cognitive fatigue, recent social interaction, location and emotional state. An objective measure of physical activity was also included in the form of a step counter function, which was realised using the smartphone's built-in step counting facility. Additionally, facial landmarks were recorded, and an objective measure of emotional state based on facial expression was determined. Accordingly, facial feature recording and analysis made use of Affectiva's Affdex SDK [64], which uses FACS [21] to analyse defined facial landmarks on the face and to then interpret them as facial expressions along with their corresponding emotions. Later offline processing also employed the facial landmarks in conjunction with average reaction time during the PVT in order to determine the role an ensemble machine learning approach might have to play in determining levels of cognitive fatigue. While realising these changes in the application's design, care was taken to retain the character of the original application's user experience design, which had been tested and validated during our previous work. Fig. 1 shows a representative selection of the tools within the application and their design.

In summary, the revised smartphone application allowed for the recording of each participant's self-reported cognitive fatigue, recent social interaction, location and emotional state, using a set of subjective self-assessments, along with an objective measure of physical activity in the form of a step counter function. Additionally, it allowed the recording of facial landmarks, facial expressions and underlying emotions using an objective facial feature analysis tool. The facility to objectively measure cognitive fatigue was also included, made possible using our previously validated set of game-like tasks, which included: PVT, Spatial Span Task and Mental Arithmetic Test.

B. STUDY PROTOCOL

The study was designed to allow individual participants to take part across a continuous two-week period, during which they would receive a daily notification to take part in the study. The timing of this notification and the subsequent completion of the daily session was customisable by the participant in order to best suit their daily routine and thus promote adherence. During each assessment session, the application first delivered a set of subjective self-assessment questions relating to cognitive fatigue, location type, recent social interaction and emotional state. The game-like tasks were then delivered to objectively assess cognitive fatigue. These were delivered in the order: Spatial Span Task, PVT and Mental Arithmetic Test. Each task was designed to be limited to a duration of 90 seconds, in keeping with our previous work. An objective measure of emotional state and facial expression was then recorded, along with a set of facial landmarks. The full assessment session lasts approximately six minutes and could be undertaken at any time and in any location that the participant felt was appropriate for them. An objective measure of physical activity was recorded in the background on a daily basis throughout the study, using a step counter function.

1) PARTICIPANT RECRUITMENT

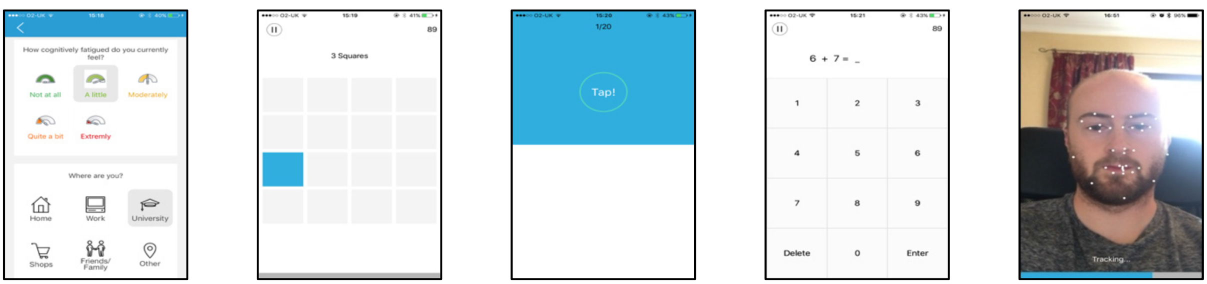
Participants (n = 28: 18 males, 10 females) were recruited within Ulster University to undertake the study over a two-week period. The mean age of participants recruited was 24 years (SD = 11) and they were all in good health. In addition, participants were required to own an iPhone to ensure they were familiar with using iOS-based smartphone applications, which removed the need for additional training in device use prior to undertaking the study. Before commencing the study, participants were individually informed of its purpose and of what was required from them. Participants were also informed that all data would be collected anonymously and that they could withdraw from the study at any time. Once the study had been verbally explained, information sheets were provided that reinforced the steps and requirements of the study and consent was taken. No inducements to participant were offered. This study received approval from the University's ethics committee (*ref: FCE 20160419 16.23*).

2) APPLICATION DEPLOYMENT

The smartphone application was installed individually on to each participant's smartphone by the principal investigator. Minimal training was required as instructional information was provided at each step within the smartphone application. Previous work had also verified the usability and visual design of the application [13], [14], [16]. Once the application was deployed to each participant's personal device, the final step was for participants to turn on notifications for the application so they could receive reminders to take part in each test session.

3) DATA COLLECTION

During each session data was recorded in relation to cognitive fatigue, social interaction, location, emotional state, physical activity and when the assessment session was performed, for later analysis against participant performance during the game-like cognitive fatigue assessment tasks. A subjective self-reported measure of cognitive fatigue was recorded using a single-item scale that asked the question: "*How fatigued do you currently feel?*", with responses on a 5-point scale recorded as: *1. Not at all, 2. A little, 3. Moderately, 4. Quite a bit,* or *5. Extremely*. An objective measure of physical activity was recorded using a step counter function that made use of pedometer data provided by iOS across a 24-hour period, running from midnight to midnight. A subjective self-reported measure of the participant's recent level of social interaction was recorded using a single-item scale that asked the question: "*How much social interaction have you had today?*", with responses on a visual scale with 11 intervals, where "*0*" indicated "None" and 10 indicated "*Lots*". A subjective self-reported location type was reported using the question: "*Where are you?*" with responses recorded as: *1. Home, 2. Work, 3. University, 4. Shops, 5. Family/Friends,*



(a) (b) (c) (d) (e)

FIGURE 1. Smartphone cognitive fatigue assessment application: (a) Cognitive fatigue self-assessment; (b) Spatial Span Task; (c) PVT; (d) Mental Arithmetic Test; (e) Facial feature analysis.

or *6. Other*. A subjective self-reported emotional state was reported using the question: "*Emotionally how do you feel?*", with responses based on Ekman's 6 emotions [65] recorded as: *1. Angry, 2. Afraid, 3. Disgusted, 4. Happy, 5. Sad,* or *6. Surprised*.

Facial feature analysis made use of the Affdex SDK from Affectiva and recorded the full set of data this tool provides. A set of 10 samples (frames) were recorded on each occasion to avoid issues relating to poor framing of the participant's face, occlusion or fleeting variations in the facial expression being exhibited which might be atypical of the participant's emotional state. For each sample, a set of 34 automatically detected facial landmarks were recorded, which were used to identify a set of 21 facial expressions. The tool recorded the probability of each of these facial expressions being present within each sample. The expressions were then used to determine which of seven emotions were present in the sample; with Contempt being an additional possible outcome over the self-reported emotions noted earlier. The tool also reported Valence and Engagement. The Emotions and Engagement were reported as a percentage of the likelihood of being present in the sample (ranging from 0% to 100%), while Valance was reported as a measure in the range [−100, 100], with negative values represented the extent of negative valence and positive valence representing the extent of any positive valence.

During the game-like cognitive fatigue assessment tasks, a set of data relating to participant performance within the games was recorded. For the Spatial Span Task, the data recorded was the: number of correctly completed sequences, number of incorrectly completed sequences, and length of the longest sequence correctly completed. For the Mental Arithmetic Test, the data recorded was the number of questions correctly answered and the number of questions incorrectly answered. For the Psychomotor Vigilance Task, the data recorded was the: number of correct (in time) reactions, response time of each correct reaction, and number of early reactions. Finally, the time at which the session was conducted was also recorded.

Data recorded during the session was temporarily stored on the device's SQLite database, in preparation for transfer via HTTPS to a secure online database. This prevented performance and data loss issues in the event of no mobile network signal being available. When a mobile network connection was available, data was automatically transferred from the internal database to the online database.

C. STATISTICAL ANALYSIS

Data collected regarding the hypothesis that: measures of a participant's physical activity, social interaction, location and emotional state would be shown to be indicative of episodes of cognitive fatigue, as measured using our established tool, were analysed statistically using SPSS version 24. Weighted two-tailed Pearson's correlation was used to test statistical significance between collected objective and subjective variables. Multiple regression was utilised to analyse relationships in cognitive testing scores. To understand the results from the categorised variable of location, descriptive statistical analysis was used. The lower p value of $<.01$ was used to determine statistical significance to eliminate the probability of a Type 1 error. On completion of the study, participants were required to fill in the System Usability Scale (SUS) [66]. This is a 10-point questionnaire with five response options for each question.

Data collected regarding the hypothesis that: feature analysis would allow for the selection of a machine learning approach capable of identifying episodes of cognitive fatigue was investigated through the implementation of a classifier ensemble, as the consistent accuracy offered from ensembles when analysing a high volume of data makes them an ideal fit for facial feature analysis. Based on the facial landmarks recorded during the study, in parallel with cognitive testing scores, the dataset utilised contained 145 facial feature sets, whereby each individual facial feature set contained 28 captured variables describing individual points of features on the face that correspond to a defined level of cognitive fatigue, as determined by validated cognitive testing methods that were simultaneously administered during the facial feature analysis [16]. Metrics obtained from these pre-validated measures were used as the classification categories. Specifically, average reaction time during the PVT was used as it has the ability to accurately measure small discrepancies in performance due to cognitive fatigue [16], [32], [33], [67]. Subsequently, the goal of the classification task was to map facial feature data into one of three groups: low, normal and high levels of cognitive fatigue. Initially, a set of popular base classifiers was chosen to be trained on the dataset including: J48 Decision Tree, Random Tree, Bayesian Network, Naive Bayes, K* Instance Based Learner, K-Nearest Neighbors, Multi-Layer Perception and Simple Logistic Regression. These individual classifiers were first assessed to determine which was the most accurate on the dataset. Each was trained on the dataset, with the highest performing classifier being chosen for implementation in classifier ensembles. Once the optimal base classifier was selected, it was employed within Bagging, Boosting, Random Forest and both Principle Component Analysis (PCA) and Random Projections (RP) based Rotation Forest ensembles. 10-fold cross validation was used to indicate the accuracy of each ensemble on an independent dataset and to eliminate overfitting of data. In addition, to evaluate the impact of ensemble size on classification accuracy, each classifier was executed 10 times with an increased ensemble size each time, thereby providing an indication of the relationship between classification accuracy and ensemble size. The quality of the classification was measured using the classification accuracy, Root Mean Square Error and the Area Under the Receiver Operating Characteristic curve.

V. RESULTS

During the two-week study, participants recorded 145 individual sessions, which represented an adherence rate of 37%.

TABLE 1. Cognitive testing results grouped by location.

Home	N	min	max	$\bar{x}$	s
Self-Assessed Fatigue	58	1	5	2.43	1.48
Self-Assessed Social Interaction	58	0	10	3.10	3.26
Spatial Span Score	58	4	9	7.14	1.37
Spatial Span Correct	58	4	6	5.24	.66
Arithmetic Questions Correct	58	6	34	28.59	4.19
Reaction Time	58	.31	.62	.44	.075
Work	**N**	**min**	**max**	**$\bar{x}$**	**s**
Self-Assessed Fatigue	19	2	5	4.37	1.17
Self-Assessed Social Interaction	19	0	10	7.05	3.19
Spatial Span Score	19	5	8	6.21	.71
Spatial Span Correct	19	5	6	5.05	.23
Arithmetic Questions Correct	19	23	29	26.53	1.58
Reaction Time	19	.44	.61	.51	.04
University	**N**	**min**	**max**	**$\bar{x}$**	**s**
Self-Assessed Fatigue	65	1	5	3.18	1.5
Self-Assessed Social Interaction	65	0	10	4.49	3.04
Spatial Span Score	65	3	9	7.06	1.38
Spatial Span Correct	65	4	6	5.29	.61
Arithmetic Questions Correct	65	23	34	28.71	2.78
Reaction Time	65	.33	.59	.44	.05

Sessions typically took six minutes to complete and 65% of sessions took place within the first hour of receiving the daily notification that was intended to prompt participation.

A. DESCRIPTIVE ANALYSIS

An initial consideration of self-reported location type showed that 143 of the 145 recorded locations fell under one of three labels: Home, Work, or University. When data from the other measures was grouped by self-reported location, as shown in Table 1, it was observed that the mean level of Self-Assessed Fatigue was reported as being highest at Work and lowest at Home, with a 57% difference between the two locations. The results from the previously validated game-like cognitive tests supported this subjective assessment, showing higher levels of cognitive fatigue at Work compared to at Home. The mean values obtained for the Spatial Span Score, Spatial Span Correct, Arithmetic Questions Correct, and Reaction Time were all lower at Work than at Home, corresponding with higher Self-Assessed Fatigue and Self-Assessed Social Interaction. Levels of Self-Assessed Social Interaction reported at Work were on average 127% higher than at Home, indicating that a higher level of social interaction has a negative effect on cognitive fatigue levels. Self-Assessed Fatigue was on average 80% higher at Work than at Home, which, again, is in agreement with data from our objective cognitive fatigue tests. Results captured at University indicated higher levels of cognitive fatigue that were 45% greater than at Home, which is, again, consistent with social interaction levels being higher at University, thus indicating social interaction as a contributing factor to cognitive fatigue.

B. WEIGHTED CORRELATION ANALYSIS

Correlation analysis was carried out on the self-assessment measures, cognitive tests and facial features that were recorded. This approach explored which measures might have a significant associative link to the results obtained from the PVT and self-assessed level of cognitive fatigue, as they are both validated baseline measures. Scores from the three cognitive tests; Spatial Span Task, the PVT and Mental Arithmetic Test, all showed a significant correlation to each other, as shown in Table 2. In addition, the tests also provided significant correlations ($p = .001$) to Self-Assessed Fatigue levels through the single-item self-assessment scale, also shown in Table 2. All cognitive testing correlations to the self-assessment scale were either above .5 or below −.5, showing the strength of all the correlations. The strongest correlation to Self-Assessed Fatigue was Average Reaction Time ($r = .643$, $p = .001$), obtained during the PVT, which is in keeping with the findings from our previous work, where PVT reaction time also displayed the most significant correlation to self-assessment using the MFS.

Higher reported levels of Self-Assessed Fatigue correlated with lower game testing scores (Average Reaction Time, Average Spatial Span Correct, Average Arithmetic Questions Correct, Average Spatial Span Score), suggesting that the cognitive tests were able to pick up on varying levels of fatigue. In Table 2, Self-Assessed Social Interaction showed significant correlation to Self-Assessed Fatigue ($r = .377$, $p = .001$), Average Spatial Span Score ($r = -.256$, $p = .002$) and Average Reaction Time ($r = .247$, $.003$). Even though all of these significant correlations are quite weak, they are still significant and because they all indicated the same thing, it can be concluded that there is a relationship between social interaction and cognitive fatigue.

Conversely, daily activity levels recorded using step count (Daily Step Count) did not show any correlation to cognitive tests or self-assessment, also shown in Table 2. As such, in this study, physical activity was not shown to have any relationship to cognitive fatigue.

TABLE 2. Pearson Product-Moment Correlations of self-reported cognitive fatigue, self-reported social interaction, step count and each cognitive test with our previously validated cognitive tests.

		Average Reaction Time	Average Spatial Span Correct	Average Arithmetic Questions Correct	Average Spatial Span Score	Self-Assessed Social Interaction
Self-Assessed Fatigue	r	.643	-.600	-.601	-.549	.377
	p	.001**	.001**	.001**	.001**	.001**
Average Reaction Time	r	1	-.498	-.544	-.448	.247
	p	–	.001**	.001**	.001**	.003
Average Spatial Span Correct	r	-.498	1	.543	.834	-.193
	p	.001**	–	.001**	.001**	.020
Average Arithmetic Questions Correct	r	-.544	.543	1	.567	-.046
	p	.001**	.001**	–	.001**	.586
Average Spatial Span Score	r	-.448	.834	.567	1	-.256
	p	.001**	.001**	.001**	–	.002
Daily Step Count	r	.051	-.063	-.080	-.063	-.011
	p	.539	.453	.341	.453	.895

**Correlation is significant at the 0.01 level (2-tailed).

TABLE 3. Pearson Product-Moment Correlations of facial feature analysis and self-reported cognitive fatigue our previously validated cognitive tests.

		Self-Assessed Fatigue	Average Spatial Span Correct	Average Spatial Span Score	Average Reaction Time
Anger	r	.257	-.084	-.152	.138
	p	.002	.316	.068	.097
Sadness	r	.199	-.088	-.126	.221
	p	.006	.293	.132	.008
Valence	r	-.271	.156	.163	-.210
	p	.001**	.062	.050	.011
Brow Furrow	r	.339	-.142	-.235	.251
	p	.001**	.089	.004	.002
Eyelid Tighten	r	.302	-.158	-.253	.171
	p	.001**	.058	.002	.039
Lip Suck	r	-.240	.184	.248	-.227
	p	.004	.027	.003	.006

**Correlation is significant at the 0.01 level (2-tailed).

Analysis of facial cues showed that there were correlations between raised levels of cognitive fatigue, as measured using our previously validated tests, and facial cues which had been classified as exhibiting Anger, Sadness, negative Valence, increased Brow Furrow, increased Eyelid Tighten, and increased Lip Suck (pulling of lips inwards to the mouth). Table 3 presents the facial analysis outputs that showed these significant correlations.

C. REGRESSION ANALYSIS

Next regression analysis was carried out on the recorded data. The *one in ten rule* is widely used to guide how many predictor parameters can be estimated from data when doing regression analysis while keeping the risk of overfitting low [68]–[70], stating that one predictive variable can be studied for every ten events. However, more recent literature has suggested relaxing this rule slightly, such that between five and nine predictor variables is acceptable [71]. Based on this, as the study had 28 participants, it was decided that three of the highest correlating variables would be used in the regression analysis. As such, Average Reaction Time, Average Spatial Span Correct and Average Arithmetic Questions Correct were entered into a multiple linear regression model, as shown in Table 4, in order to determine which were predictive of the dependent variable:

TABLE 4. Standardized and unstandardized regression coefficients regression model.

	Unstandardized Coefficients		Standardized Coefficients*		
	B	*SE*	*β*	*T*	*95% CI*
Average Reaction Time	8.741	1.664	.368	5.254	(5.452, 12.031)
Average Spatial Span Correct	-.738	.182	-.283	-4.044	(-1.098, -.377)
Average Arithmetic Questions Correct	-.111	.033	-.247	-3.418	(-.176, -.047)

Dependent variable: Self-Assessed Fatigue
*Significance at the 0.01 level shown for all variables tested

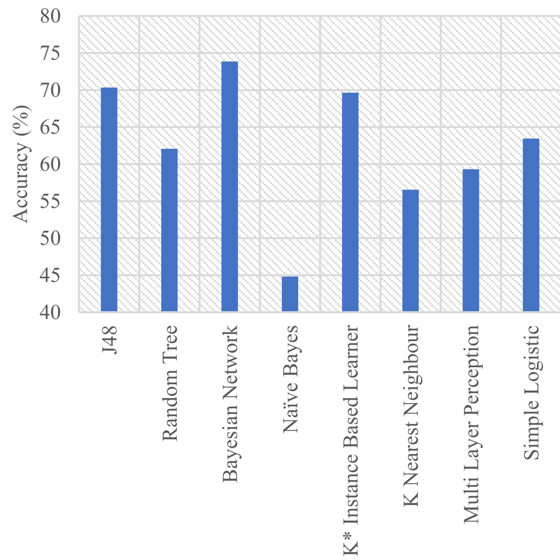


FIGURE 2. Comparison of individual classifier performance.

Self-Assessed Fatigue. A statistically significant model was shown (p = .001, F = 45.333), which accounted for 55% (adjusted R square .552) of the variance observed, with all variables having made a significant contribution to the overall model.

D. ENSEMBLE CLASSIFIER ANALYSIS

Individual classification accuracies obtained from each of the base classifiers investigated for the offline classification of ternary levels of cognitive fatigue, based on PVT average reaction time from facial feature data, are shown in Fig. 2. As may be observed, the use of a Bayesian Network obtained the highest classification accuracy (73.86%), which was marginally superior to the use of a J48 Decision Tree (70.34%). Consequently, the Bayesian Network was employed as the base classifier during comparative analysis of the ensemble classifiers.

Employing the Bayesian Network as the base classifier, a comparison of the classification accuracies obtained for differing sizes of a variety of ensemble classifiers, is given in Table 5, with the corresponding Root Mean Square Error (RMSE) values obtained given in Table 6.

From Table 5 it can be observed that the performance of a Rotation Forest ensemble classifier can vary greatly depending on the transformation method used, however, the use of PCA with Rotation Forest outperformed all other ensemble classifiers, obtaining the highest classification accuracy (82.17%) for an ensemble size of 90. With regard to the RMSE values obtained, Table 6 shows the PCA-based Rotation Forest ensemble achieved the largest number of low RMSE values across the range of classifier sizes tested, with the lowest RMSE value (0.34) occurring when the ensemble size was 80 or 100.

While the RMSE values obtained were very similar for both implementations of Rotation Forest ensemble, the PCA-based variation has been shown to consistently produce lower RMSE values than those obtained from the RP-based variation. To further evaluate the impact of ensemble size and feature selection on both variations of Rotation Forest ensemble, the corresponding Area Under the Receiver Operating Characteristic curve (AUC) values were determined, as given in Table 7.

From Table 7 it may be observed that the use of PCA with Rotation Forest achieves the most promising results, achieving the highest AUC value for five different ensemble sizes, thereby indicating the PCA-based Rotation Forest ensemble produces a larger true positive rate, with the highest AUC value occurring for an ensemble size of 90. Consequently, given the brief evaluation of the set of classifiers, the offline use of a Bayesian Network-based ensemble, employing PCA and an ensemble size of 90, potentially achieves a high classification accuracy (82.17%) in distinguishing between low, normal and high levels of cognitive fatigue from facial feature data.

E. PARTICIPANT FEEDBACK

After the two-week study period was completed, each participant was asked to complete user feedback, via the SUS, and openly comment on any aspect of the application they found engaging or unlikeable. Of the 28 participants, 14 completed the follow-up questionnaire.

Accordingly, the smartphone application received an average SUS score of 86/100 (SD = 14), thereby indicating an excellent, above average usability rating overall. The primary reason reported for non-adherence to the daily test was due to the repetitive nature of the tasks. It was suggested that variation in the tests used within the application would improve the application and help promote participation.

TABLE 5. Classifier accuracy for varying ensemble sizes (highest value from each classifier highlighted).

	Ensemble Size									
Ensemble Type	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
Random Forest	77.79	**80.93**	78.48	79.17	78.48	78.48	78.48	79.17	80.55	80.55
Rotation Forest (PCA)	78.03	81.47	80.17	80.55	80.48	81.17	81.17	81.55	**82.17**	81.17
Rotation Forest (RP)	76.41	75.72	77.79	77.79	77.79	78.48	**80.55**	79.86	79.86	78.48
Bagging	77.10	77.10	77.79	77.79	77.79	77.79	77.10	**78.48**	77.79	77.79
Boosting	73.66	75.03	76.41	79.17	79.86	81.14	80.74	81.24	**81.93**	81.24

TABLE 6. RMSE value of varying ensemble size (lowest value highlighted in each column).

	Ensemble Size									
Ensemble Type	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
Random Forest	**0.35**	**0.35**	0.36	0.36	0.36	0.36	0.36	0.35	**0.35**	0.35
Rotation Forest (PCA)	0.36	0.36	**0.35**	**0.35**	**0.35**	**0.35**	**0.35**	**0.34**	**0.35**	**0.34**
Rotation Forest (RP)	0.37	0.36	0.36	0.36	0.36	0.36	**0.35**	0.35	**0.35**	0.35
Bagging	0.38	0.36	0.37	0.37	0.37	0.37	0.36	0.36	0.36	0.36
Boosting	0.43	0.43	0.41	0.40	0.39	0.39	0.39	0.38	0.38	0.38

TABLE 7. AUC values for Rotation Forest ensemble classifiers of varying ensemble size (highest values highlighted).

	Ensemble Size									
Ensemble Type	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
Rotation Forest (PCA)	0.86	0.86	**0.88**	**0.88**	**0.88**	**0.88**	**0.89**	**0.88**	**0.90**	**0.89**
Rotation Forest (RP)	0.86	0.86	0.87	0.87	**0.88**	0.878	0.88	**0.88**	0.88	0.88

VI. DISCUSSION AND FUTURE WORK

As hypothesised, all the cognitive tests showed a degradation in participant performance during higher levels of self-reported fatigue. Low participation rates could be considered as a drawback to the current approach and, as such, further work to improve adherence should be conducted. Randomisation of the tests that are presented to participants may reduce the repetitive nature of the application while simultaneously reducing the time on task required during each session. As indicated from testing results, assessment of cognitive fatigue is feasible via a smartphone application, however, adherence to regular testing is crucial to gain a longitudinal understanding of the condition. Participant feedback also suggested that more detailed scoring and feedback during cognitive tests would increase the competitive aspect of the application, which, in turn, could potentially increase daily participation. A variation in tasks presented may also promote adherence. For example, the Spatial Span Task measured a participant's working memory, therefore this could be substituted for another cognitive test that also measures this aspect, such as N-Back [46], Digit Span [72] or Paired Associates [73].

As has been observed in the self-reported locations, where sessions were completed, the majority took place at either Home, Work or University. This is potentially due to the recruitment of participants within a university setting. Future work will aim to employ a more diverse cohort to explore this aspect further. However, from the current results obtained, it can be observed that cognitive fatigue levels are generally lower when at Home in comparison with other locations. Accordingly, this knowledge could be employed on an individual basis to help tailor the location a test is delivered, which would permit targeted assessment and intervention during periods of potentially higher cognitive fatigue.

Prior to this study it was considered that the collection of physical activity data would help identify instances of cognitive fatigue. Through analysis of the inbuilt recorded physical activity results obtained it was observed that there was no correlation between the number of steps taken per day with any other metric collected during a session. This may be due to the nature of the physical activity that was carried out, such that it was not excessive beyond normal everyday activity, therefore did not have any significant impact on cognitive fatigue levels. Logically this may indicate that normal daily activity does not contribute to cognitive fatigue. Future work could include additional self-assessment that asks if recent physical activity was excessive for the participant based on normal physical activity encountered.

It was also hypothesised that facial feature analysis would indicate specific traits that are present during periods of higher cognitive fatigue. Facial detection carried out during each session showed dominant emotions during higher levels of cognitive fatigue, specifically anger and sadness. Facial feature analysis also revealed that there were several facial cues that were regularly dominant during higher cognitive fatigue, specifically negative valence, increased brow furrow, increased eyelid tightening and increased lip suck. Consequently, future work will investigate the combination of testing scores and facial features to detect the onset of cognitive fatigue; the inclusion of additional measures within the application has shown that facial feature analysis and location information can assist in the assessment of cognitive fatigue when used alongside subjective and objective measures. Utilising a combination of these measures within a smartphone application may provide the opportunity to deliver feedback though a smartphone rather than an intervention being dependent upon a clinical environment.

Investigations into the offline classification of low, normal and high levels of cognitive fatigue from facial feature landmarks indicated that a PCA-based Rotation Forest ensemble using a Bayesian Network as the base classifier outperforms all other ensemble classifiers tested over a range of ensemble sizes. The highest classification accuracy was obtained when the ensemble size is 90, with the RMSE values obtained indicating the Rotation Forest ensemble requires a larger ensemble size in order to generate a more diverse classifier. Correspondingly, the AUC values obtained further indicate that the PCA-based Rotation Forest is the most successful of the ensemble classifiers tested for this classification problem; overall performance, as evaluated through Accuracy, RMSE and AUC revealed the PCA-based Rotation Forest outperforms the other ensemble classifiers tested. However, one of the main constraints of the ensemble size is the computational effort required to generate and use the classification model. Although an offline approach to machine learning was employed for the purposes of the study reported herein, the integration of modern cloud-based machine learning frameworks would potentially facilitate the generation and use of an online classifier to accompany the smartphone application, hence further supporting early detection of cognitive fatigue in-situ. Consequently, future work will investigate the use of online machine learning approaches for the development and integration of a classification model, along with further optimisation of the feature space based on the facial feature analysis to potentially reduce the necessity for large, computationally expensive ensemble approaches.

As observed from the feedback collected through the SUS, there was an increase in the SUS score obtained compared to the score of 73 obtained during previous work [16], where the application also employed the MFS self-assessment questionnaire. Replacement of the MFS questionnaire, which was previously reported by participants as reducing engagement, has potentially increased the application's usability. In addition, results obtained from the current evaluation indicated that the primary reason for non-adherence was the repetitive nature of the cognitive tests over time.

A. LIMITATIONS

Recruitment for this study took place primarily within a student population from Ulster University. A wider variety of participants with a more varying level of age and education may have been beneficial to gain a wider insight to cognitive fatigue in a more balanced population. Furthermore, the population employed did not have any previously diagnosed cognitive fatigue or cognitive deficiencies, therefore, future work should address this by recruiting from a clinical population. Accordingly, testing with a clinical cohort would potentially support stronger claims for the in-situ assessment capabilities of the smartphone application. Correspondingly, an additional limitation of this work is size of the participant group that was used for the study; increasing participant numbers would further bolster the statistical findings. Furthermore, adherence during the study was relatively low (37%), hence future work should aim to increase the level of adherence, not only to facilitate increased validation of the proposed smartphone method, but also to ensure there are no usability factors influencing adherence. Due to the application being limited to the iOS platform, deployment was restricted to participants that had access to an iOS-based device.

VII. CONCLUSION

Building upon previous work, this study highlighted smartphone evaluation methods that are accurate and practical for in-situ smartphone assessment of cognitive fatigue. Consequently, the proposed approach could supply clinicians with valuable additional information that can help in their understanding and evaluation of the condition. Both objective and subjective self-assessment have been shown to be viable in the assessment of cognitive fatigue using a smartphone, however, further research needs to be carried out with a larger population to ensure adherence rates do not negatively affect evaluation of cognitive fatigue. Increasing the length of the overall study that was carried out would enable further data collection, increasing the statistical reliability of the study.

A consideration of facial feature analysis has indicated several features that are present during higher levels of cognitive fatigue. As such, future work may explore this aspect to investigate if there are definitive features that can be selected to detect cognitive fatigue and if cognitive testing methods could be used to supplement this. The inclusion of additional measures, namely facial feature and location analysis have provided a more detailed analysis into cognitive fatigue evaluation on a smartphone. Investigations conducted into classification of levels of fatigue from facial landmarks shows the potential of incorporating such machine learning analysis; the single-camera, mobile-centered collection of data discussed herein achieved an accuracy of 82.2%, hence may be considered as an effective approach for cognitive fatigue detection. Incorporating such training and deployment of a classification model into an online approach could potentially enable the creation of a more personalised and pervasive cognitive fatigue detection application. By contrast, physical activity levels did not show any relationship to the validated approaches and more work may be needed to identify if there is a significant relationship with these metrics. The findings presented in this research indicate the effectiveness of using a smartphone application to measure and assess cognitive fatigue in-situ. Building upon previous work, the incorporation of cognitive, social, physical and emotional measures has increased the accuracy of assessment by allowing a broader range of metrics to be evaluated.

REFERENCES

[1] K. Herlofson and J. P. Larsen, "Measuring fatigue in patients with Parkinson's disease—The fatigue severity scale," *Eur. J. Neurol.*, vol. 9, no. 6, pp. 595–600, 2002.

[2] B. Johansson and L. Rönnbäck, "Mental fatigue and cognitive impairment after an almost neurological recovered stroke," *ISRN Psychiatry*, vol. 2012, pp. 1–7, Jun. 2012.

[3] L. Whitehead, "The measurement of fatigue in chronic illness: A systematic review of unidimensional and multidimensional fatigue measures," *J. Pain Symptom Manage.*, vol. 37, no. 1, pp. 107–128, 2009.

[4] P. Alhola and P. Polo-Kantola, "Sleep deprivation: Impact on cognitive performance," *Neuropsychiatric Disease Treat.*, vol. 3, no. 5, pp. 553–567, 2007.

[5] D. R. van Langenberg and P. R. Gibson, "Factors associated with physical and cognitive fatigue in patients with crohn's disease: A cross-sectional and longitudinal study," *Inflammatory Bowel Diseases*, vol. 20, no. 1, pp. 115–125, Jan. 2014.

[6] L. S. Aaronson, C. S. Teel, V. Cassmeyer, G. B. Neuberger, L. Pallikkathayil, J. Pierce, and A. N. Press, "Defining and measuring fatigue," *J. Nursing Scholarship*, vol. 31, no. 1, pp. 45–50, 1999.

[7] G. B. Neuberger, "Measures of fatigue: The fatigue questionnaire, fatigue severity scale, multidimensional assessment of fatigue scale, and short form-36 vitality (energy/fatigue) subscale of the short form health survey," *Arthritis Care Res., Off. J. Amer. College Rheumatology*, vol. 49, no. S5, pp. S175–S183, Oct. 2003.

[8] E. D. Dolan, D. Mohr, M. Lempa, S. Joos, S. D. Fihn, K. M. Nelson, and C. D. Helfrich, "Using a single item to measure burnout in primary care staff: A psychometric evaluation," *J. Gen. Internal Med.*, vol. 30, no. 5, pp. 582–587, 2015.

[9] M. L. M. van Ooff, S. A. E. Geurts, M. A. J. Kompier, and T. W. Taris, "'How fatigued do you currently feel?' Convergent and discriminant validity of a single-item fatigue measure," *J. Occupational Health*, vol. 49, no. 3, pp. 224–234, 2007.

[10] L. Bergkvist, "Appropriate use of single-item measures is here to stay," *Marketing Lett.*, vol. 26, no. 3, pp. 245–255, 2015.

[11] Z. Butt, L. I. Wagner, J. L. Beaumont, J. A. Paice, A. H. Peterman, D. Shevrin, J. H. V. Roenn, G. Carro, J. L. Straus, J. C. Muir, and D. Cella, "Use of a single-item screening tool to detect clinically significant fatigue, pain, distress, and anorexia in ambulatory cancer practice," *J. Pain Symptom Manage.*, vol. 35, no. 1, pp. 20–30, 2008.

[12] V. E. Dahiphale and S. R. Rao, "A review paper on portable driver monitoring system for real time fatigue," in *Proc. 1st Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Feb. 2015, pp. 558–560.

[13] W. Tu, L. Wei, W. Hu, Z. Sheng, H. Nicanfar, X. Hu, E. C.-H. Ngai, and V. C. M. Leung, "A survey on mobile sensing based mood-fatigue detection for drivers," in *Proc. Int. Summit, Smart City 360°*, 2016, pp. 3–15.

[14] C. Timmers, A. Maeghs, M. Vestjens, C. Bonnemayer, H. Hamers, and A. Blokland, "Ambulant cognitive assessment using a smartphone," *Appl. Neuropsychol. Adult*, vol. 21, no. 2, pp. 136–142, 2014.

[15] R. Kawamura, N. Takemura, and K. Sato, "Mental fatigue estimation based on facial expressions during speech," in *Proc. IEEE/SICE Int. Symp. Syst. Integr.*, Dec. 2015, pp. 223–228.

[16] E. Price, G. Moore, L. Galway, and M. Linden, "Validation of a smartphone-based approach to *in situ* cognitive fatigue assessment," *JMIR mHealth uHealth*, vol. 5, no. 8, p. e125, Aug. 2017.

[17] E. Price, G. Moore, L. Galway, and M. Linden, "From paper to play-design and validation of a smartphone based cognitive fatigue assessment application," in *Proc. 10th Int. Conf. Ubiquitous Comput. Ambient Intell. (UCAmI)*, in Lecture Notes in Computer Science, vol. 10069. Gran Canaria, Spain: San Bartolomé de Tirajana, Nov./Dec. 2016, pp. 321–332.

[18] E. Price, G. Moore, L. Galway, and M. Linden, "Towards a mobile assistive technology for monitoring and assessing cognitive fatigue in individuals with acquired brain injury," in *Proc. IEEE Int. Conf. Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Autonomic Secure Comput. Pervasive Intell. Comput.*, no. 15, Oct. 2015, pp. 1487–1491.

[19] E. Price, G. Moore, L. Galway, and M. Linden, "User centred design of a smartphone-based cognitive fatigue assessment application," in *Proc. 14th Int. Conf. Adv. Mobile Comput. Multi Media*, 2016, pp. 120–127.

[20] J. R. Price, E. Mitchell, E. Tidy, and V. Hunot, "Cognitive behaviour therapy for chronic fatigue syndrome in adults," *Cochrane Database Syst. Rev.*, vol. 3, Jan. 2008, Art. no. CD001027.

[21] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system (FACS)," Tech. Meas. Facial Action Consultancy, Palo Alto, CA, USA, Tech. Rep., 1978, vol. 22.

[22] B. Johansson, A. Starmark, P. Berglund, M. Rödholm, and L. Rönnbäck, "A self-assessment questionnaire for mental fatigue and related symptoms after neurological disorders and injuries," *Brain Injury*, vol. 24, no. 1, pp. 2–12, Jan. 2010.

[23] L. B. Krupp, N. G. LaRocca, J. Muir-Nash, and A. D. Steinberg, "The fatigue severity scale: Application to patients with multiple sclerosis and systemic lupus erythematosus," *Arch. Neurol.*, vol. 46, no. 10, pp. 1121–1123, 1989.

[24] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, "Visual analogue scale to evaluate fatigue severity (VAS-F)," in *STOP, THAT and One Hundred Other Sleep Scales*. New York, NY, USA: Springer, 2012, pp. 399–402.

[25] B. Johansson and L. Rönnbäck, "Mental fatigue; a common long term consequence after a brain injury," in *Brain Injury: Functional Aspects, Rehabilitation and Prevention*. Rijeka, Croatia, 2012.

[26] M. Rödholm, J.-E. Starmark, E. Svensson, and C. Von Essen, "Astheno-emotional disorder after aneurysmal SAH: Reliability, symptomatology and relation to outcome," *Acta Neurologica Scandmavica*, vol. 103, no. 6, pp. 379–385, 2001.

[27] K. A. Lee, G. Hicks, and G. Nino-Murcia, "Validity and reliability of a scale to assess fatigue," *Psychiatry Res.*, vol. 36, no. 3, pp. 291–298, 1991.

[28] A. G. E. M. de Boer, J. J. B. van Lanschot, P. F. M. Stalmeier, J. W. van Sandick, J. B. F. Hulscher, J. C. J. M. de Haes, and M. A. G. Sprangers, "Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life?" *Qual. Life Res.*, vol. 13, no. 2, pp. 311–320, 2004.

[29] B. Johansson, P. Berglund, and L. Rönnbäck, "Mental fatigue and impaired information processing after mild and moderate traumatic brain injury," *Brain Injury*, vol. 23, nos. 13–14, pp. 1027–1040, Dec. 2009.

[30] M. Kay, K. Rector, S. Consolvo, B. Greenstein, J. O. Wobbrock, N. F. Watson, and J. A. Kientz, "PVT-touch: Adapting a reaction time test for touchscreen devices," in *Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthcare*, 2013, pp. 248–251.

[31] H. P. A. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges, "The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation," *Sleep*, vol. 26, no. 2, pp. 117–126, 2003.
[32] S. A. Ferguson, B. P. Smith, M. Browne, and M. J. Rockloff, "Fatigue in emergency services operations: Assessment of the optimal objective and subjective measures using a simulated wildfire deployment," *Int. J. Environ. Res. Public Health*, vol. 13, no. 2, p. 171, 2016.
[33] K. Liu, B. Li, S. Qian, Q. Jiang, L. Li, W. Wang, G. Zhang, Y. Sun, and G. Sun, "Mental fatigue after mild traumatic brain injury: A 3D-ASL perfusion study," *Brain Imag. Behav.*, vol. 10, no. 3, pp. 857–868, 2016.
[34] D. A. Grant, K. A. Honn, M. E. Layton, S. M. Riedy, and H. P. A. Van Dongen, "3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation," *Behav. Res. Methods*, vol. 49, no. 3, pp. 1020–1029, 2016.
[35] D. Wechsler, *Wechsler Adult Intelligence Scale-Fourth Edition: Administration and Scoring Manual*. San Antonio, TX, USA: Psychological Corporation, 2008.
[36] R. M. Reitan and D. Wolfson, *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*, vol. 4. Tucson, Arizona: Reitan Neuropsychology, 1985.
[37] J. Patterson, "F-A-S test," in *Encyclopedia of Clinical Neuropsychology*. New York, NY, USA: Springer, 2011, pp. 1024–1026.
[38] B. Jarpsten and K. Taube, "DLS reading speed and spelling ability," Dept. Psychol., Stockholm Univ., Stockholm, Sweden, Tech. Rep., 1997.
[39] M. Allard, M. Husky, G. Catheline, A. Pelletier, B. Dilharreguy, H. Amieva, K. Pérès, A. Foubert-Samier, J.-F. Dartigues, and J. Swendsen, "Mobile technologies in the early detection of cognitive decline," *PLoS ONE*, vol. 9, no. 12, 2014, Art. no. e112197.
[40] M. J. Morón, R. Yáñez, D. Cascado, C. Suárez-Mejías, and J. L. Sevillano, "A mobile memory game for patients with acquired brain damage: A preliminary usability study," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Jun. 2014, pp. 302–305.
[41] J. Batista, "A drowsiness and point of attention monitoring system for driver vigilance," in *Proc. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Sep./Oct. 2007, pp. 702–708.
[42] D. Gartenberg, R. McGarry, D. Pfannenstiel, D. Cisler, T. Shaw, and R. Parasuraman, "Development of a neuroergonomic application to evaluate arousal," in *Advances in Cognitive Engineering and Neuroergonomics*. Jul. 2012, pp. 239–248. doi: 10.1201/b12313.
[43] D. Swendeman, W. S. Comulada, N. Ramanathan, M. Lazar, and D. Estrin, "Reliability and validity of daily self-monitoring by smartphone application for health-related quality-of-life, antiretroviral adherence, substance use, and sexual behaviors among people living with HIV," *AIDS Behav.*, vol. 19, no. 2, pp. 330–340, 2015.
[44] D. Kos, J. Raeymaekers, A. Van Remoortel, M. D'hooghe, G. Nagels, M. D'Haeseleer, E. Peeters, T. Dams, and T. Peeters, "Electronic visual analogue scales for pain, fatigue, anxiety and quality of life in people with multiple sclerosis using smartphone and tablet: A reliability and feasibility study," *Clin. Rehabil.*, vol. 31, no. 9, pp. 1215–1225, 2017.
[45] T. Archer, "Influence of physical exercise on traumatic brain injury deficits: Scaffolding effect," *Neurotoxicity Res.*, vol. 21, no. 4, pp. 418–434, 2012.
[46] C. L. Hogan, J. Mata, and L. L. Carstensen, "Exercise holds immediate benefits for affect and cognition in younger and older adults," *Psychol. Aging*, vol. 28, no. 2, pp. 587–594, 2013.
[47] E. K. Wise, J. M. Hoffman, J. M. Powell, C. H. Bombardier, and K. R. Bell, "Benefits of exercise maintenance after traumatic brain injury," *Arch. Phys. Med. Rehabil.*, vol. 93, no. 8, pp. 1319–1323, 2012.
[48] S. J. Blondell, R. Hammersley-Mather, and J. L. Veerman, "Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies," *BMC Public Health*, vol. 14, no. 1, 2014, Art. no. 510.
[49] D. Laurin, R. Verreault, J. Lindsay, K. MacPherson, and K. Rockwood, "Physical activity and risk of cognitive impairment and dementia in elderly persons," *Arch. Neurol.*, vol. 58, no. 3, pp. 498–504, 2001.
[50] B. D. Castell, N. Kazantzis, and R. E. Moss-Morris, "Cognitive behavioral therapy and graded exercise for chronic fatigue syndrome: A meta-analysis," *Clin. Psychol. Sci. Pract.*, vol. 18, no. 4, pp. 311–324, 2011.
[51] L. Gan, B. Cui, and W. Wang, "Driver fatigue detection based on eye tracking," in *Proc. 6th World Congr. Intell. Control Autom.*, vol. 2, Jun. 2006, pp. 5341–5344.
[52] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imag.*, vol. 8, no. 5, pp. 357–377, 2002.
[53] J.-J. Yan, H.-H. Kuo, Y.-F. Lin, and T.-L. Liao, "Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing," in *Proc. IEEE Int. Symp. Comput. Consum. Control (ISC)*, Jun. 2016, pp. 243–246.
[54] N. Edenborough, R. Hammoud, A. Harbach, A. Ingold, B. Kisacanin, P. Malawey, T. Newman, G. Scharenbroch, S. Skiver, M. Smith, A. Wilhelm, G. Witt, E. Yoder, and H. Zhang, "Driver state monitor from DELPHI," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 1–2.
[55] M. M. Bundele and R. Banerjee, "An SVM classifier for fatigue-detection using skin conductance for use in the bits-lifeguard wearable computing system," in *Proc. 2nd Int. Conf. Emerg. Trends Eng. Technol.*, Dec. 2009, pp. 934–939.
[56] H. Al-Libawy, A. Al-Ataby, W. Al-Nuaimy, M. A. Al-Taee, and Q. Al-Jubouri, "Fatigue detection method based on smartphone text entry performance metrics," in *Proc. 9th Int. Conf. Develop. eSyst. Eng. (DeSE)*, Aug./Sep. 2016, pp. 40–44.
[57] D. Ghimire and J. Lee, "Extreme learning machine ensemble using bagging for facial expression recognition," *J. Inf. Process. Syst.*, vol. 10, no. 3, pp. 443–458, 2014.
[58] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 886–893.
[59] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Gool, "Random forests for real time 3D face analysis," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 437–458, 2013.
[60] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[61] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, pp. 119–139, Aug. 1997.
[62] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[63] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
[64] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. el Kaliouby, "AFFDEX SDK: A cross-platform real-time multi-face expression recognition toolkit," in *Proc. CHI Conf. Extended Abstracts Hum. Factors Comput. Syst.*, May 2016, pp. 3723–3726.
[65] P. Ekman, "An argument for basic emotions," *Cognit. Emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992.
[66] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
[67] S. P. A. Drummond, A. Bischoff-Grethe, D. F. Dinges, L. Ayalon, S. C. Mednick, and M. J. Meloy, "The neural basis of the psychomotor vigilance task," *Sleep*, vol. 28, no. 9, pp. 1059–1068, 2005.
[68] F. E. Harrell, Jr., K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati, "Regression modelling strategies for improved prognostic prediction," *Statist. Med.*, vol. 3, no. 2, pp. 143–152, Apr. 1984.
[69] F. E. Harrell, K. L. Lee, and D. B. Mark, "Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statist. Med.*, vol. 15, no. 4, pp. 361–387, 1996.
[70] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *J. Clin. Epidemiol.*, vol. 49, no. 12, pp. 1373–1379, Dec. 1996.
[71] E. Vittinghoff and C. E. McCulloch, "Relaxing the rule of ten events per variable in logistic and Cox regression," *Amer. J. Epidemiol.*, vol. 165, no. 6, pp. 710–718, Jan. 2007.
[72] B. Johansson and L. Ronnback, "Mental fatigue scale and its relation to cognitive, social and emotional functioning after a TBI or stroke," *Brain Injury*, vol. 28, no. 1, pp. 572–573, 2014.
[73] S. H. A. Chen, J. D. Thomas, R. L. Glueckauf, and O. L. Bracy, "The effectiveness of computer-assisted cognitive rehabilitation for persons with traumatic brain injury," *Brain Injury*, vol. 11, no. 3, pp. 197–210, 1997.

EDWARD PRICE received the B.Sc. degree (Hons.) in interactive multimedia design and the Ph.D. degree in smart healthcare from the School of Computing, Ulster University, U.K., in 2014 and 2019, respectively.

GEORGE MOORE received the B.Sc. degree (Hons.) in computing and the Ph.D. degree in image processing from Ulster University, U.K., in 1994 and 2000, respectively.

Since 2000, he has been a Member of Academic Staff with the School of Computing, Ulster University, first as a Lecturer and then as a Senior Lecturer, where he is currently a member of the Pervasive Computing Research Group. His current research interests include: human–computer interaction, affective computing, positive computing, mobile technologies, machine learning, and smart healthcare.

LEO GALWAY received the B.Sc. degree in computer science from Queen's University Belfast, U.K., in 1998, and the M.Sc. degree in computing and intelligent systems and the Ph.D. degree in machine learning for digital games from Ulster University, U.K., in 2004 and 2009, respectively.

In 2010, he joined the School of Computing, Ulster University, as a Research Associate, where he was appointed as a Lecturer in computer science, in 2012, and he specializes in teaching mobile technologies. His research interests include affective computing, brain–computer interfaces, machine learning, and ubiquitous computing.

MARK LINDEN received the B.A. degree (Hons.) in psychology and the Ph.D. degree in environmental psychology from Queen's University Belfast, U.K., in 1995 and 2000, respectively, and the M.Sc. degree in cognitive neuropsychology from the University of London, U.K., in 2007.

In 2006, he was a Lecturer with the School of Nursing and Midwifery, Queen's University Belfast, where he is currently a Senior Lecturer in health sciences. He is also a member of the Maternal and Child Health Research Team, School of Nursing and Midwifery, and a Fellow of the Centre for Evidence and Social Innovation, Queen's University Belfast. His research interests include the social and cognitive impact of paediatric brain injury on the child, family, and the social world.